NPSCS-92-011

NAVAL POSTGRADUATE SCHOOL
Monterey, California

Exploiting Captions in Retrieval of Multimedia Data

Neil C. Rowe
Eugene J. Guglielmo

July 1992

Approved for public release; distribution is unlimited.

Prepared for:

Naval Postgraduate School
Monterey, California 93943

FedDocs
D 208.14/2
NPS-CS-92-011

NAVAL POSTGRADUATE SCHOOL

Monterey, California

REAR ADMIRAL R. W. WEST, JR.
Superintendent

HARRISON SHULL
Provost

This report was prepared for and funded by the Naval Postgraduate School, Monterey, California.

Reproduction of all or part of this report is authorized.

This report was prepared by:

VALDIS BERZINS
Associate Chairman for
Technical Research

PAUL MARTO
Dean of Research

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUA
MONTEREY CA 939

REPORT DOCUMENTATION PAGE

1a. REPORT SECURITY CLASSIFICATION UNCLASSIFIED	1b. RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3. DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited
2b. DECLASSIFICATION/DOWNGRADING SCHEDULE	
4. PERFORMING ORGANIZATION REPORT NUMBER(S) NPSCS-92-011	5. MONITORING ORGANIZATION REPORT NUMBER(S)
6a. NAME OF PERFORMING ORGANIZATION Computer Science Dept. Naval Postgraduate School 6b. OFFICE SYMBOL (if applicable) CS	7a NAME OF MONITORING ORGANIZATION Naval Ocean Systems Center
6c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943	7b. ADDRESS (City, State, and ZIP Code) San Diego, CA 92152
8a. NAME OF FUNDING/SPONSORING ORGANIZATION Naval Postgraduate School 8b. OFFICE SYMBOL (if applicable) NPS	9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER O&MN Direct Funding
8c. ADDRESS (City, State, and ZIP Code) Monterey, CA 93943	10 SOURCE OF FUNDING NUMBERS PROGRAM ELEMENT NO. / PROJECT NO. / TASK NO / WORK UNIT ACCESSION NO.

11. TITLE (Include Security Classification)
Exploiting Captions in Retrieval of Multimedia Data

12. PERSONAL AUTHOR(S)
Neil C. Rowe and Eugene J. Guglielmo

13a. TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
	FROM 10/91 TO 9/92	July 1992	24

16. SUPPLEMENTARY NOTATION

17. COSATI CODES			18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
FIELD	GROUP	SUB-GROUP	Databases, Natural-language, Captions, Multimedia

19. ABSTRACT (Continue on reverse if necessary and identify by block number)

Descriptive natural-language captions can help organize multimedia data. We described our MARIE system that interprets English queries directing the fetch of media objects. it is novel in the extent to which it exploits previously interpreted and indexed English captions for the media objects. Our routine filtering of queries through descriptively-complex captions (as opposed to keyword lists) before retrieving data can actually improve retrieval speed, as media data are often bulky and time consuming to retrieve, difficult upon which to perform content analysis, and even small improvements to query prevision can often pay off. Handling the English of captions and queries about them is not as difficult as it might seems, as the matching does not require deep understanding, just a comprehensive type hierarchy for caption concepts. An important innovation of MARIE is "supercaptions" describing sets of captions, which can minimize caption redundancy.

20. DISTRIBUTION/AVAILABILITY OF ABSTRACT ☒ UNCLASSIFIED/UNLIMITED ☐ SAME AS RPT. ☐ DTIC USERS	21. ABSTRACT SECURITY CLASSIFICATION UNCLASSIFIED
22a. NAME OF RESPONSIBLE INDIVIDUAL Neil C. Rowe	22b. TELEPHONE (Include Area Code) (408) 646-2462 22c. OFFICE SYMBOL CSRp

Exploiting Captions in Retrieval of Multimedia Data

Neil C. Rowe and Eugene J. Guglielmo[1]

Department of Computer Science

Code CS/Rp, U. S. Naval Postgraduate School

Monterey, CA USA 93943

(rowe@cs.nps.navy.mil)

ABSTRACT

Descriptive natural-language captions can help organize multimedia data. We describe our MARIE system that interprets English queries directing the fetch of media objects. It is novel in the extent to which it exploits previously interpreted and indexed English captions for the media objects. Our routine filtering of queries through descriptively-complex captions (as opposed to keyword lists) before retrieving data can actually improve retrieval speed, as media data are often bulky and time-consuming to retrieve, difficult upon which to perform content analysis, and even small improvements to query precision can often pay off. Handling the English of captions and queries about them is not as difficult as it might seem, as the matching does not require deep understanding, just a comprehensive type hierarchy for caption concepts. An important innovation of MARIE is "supercaptions" describing sets of captions, which can minimize caption redundancy.

[1] This work was sponsored by the Naval Ocean Systems Center in San Diego, California, the Naval Air Warfare Center in China Lake, California, and the U. S. Naval Postgraduate School under funds provided by the Chief for Naval Operations.

1. Introduction

Captions have historically been an essential tool in organizing and accessing multimedia data, especially nontextual data. Captions in natural language can embody the classificatory information and heuristic advice necessary to navigate through very large data collections. Unfortunately, no current database systems exploit natural-language captions in a comprehensive way for data access. Many multimedia database systems store text information, but most just store it as another data item that cannot help retrieve related data items. Some systems, such as the existing one for the Photo Lab at the Naval Air Warfare Center in China Lake, CA, USA, index multimedia data from isolated keywords extracted from captions, ignoring valuable information present in the caption. For instance:

> Within the strands of the wire coral forest, schools of three-inch-long cardinal fish hover facing into the current, their silvery skins mirroring the camera's electronic flash. *(National Geographic, Oct. 1990, p. 22)*

If we index this caption on its principal keywords "coral," "forest," "schools," "cardinal," "fish," "current," "skins," "camera," and "flash," we can get false hits in querying "cardinals in forests," "fish in high schools," and "cameras with low current electronic flashes." We could prefer the matches that match more words of the query, but this does not prevent the fundamental misunderstandings in the three matches. Some work in information retrieval has linked nouns to corresponding adjectives for keyword lookup, but this handles only part of the problem, and what is clearly needed is a full parse and semantic interpretation of captions and queries using methods of language understanding and knowledge representation from artificial intelligence. Full natural language descriptions would avoid most ambiguity problems of words in keyword lists, improving the query match precision.

General natural-language understanding remains an unsolved problem, but handling captions and queries about them is much simpler for four reasons. First, full understanding is not necessary to retrieve data. For instance, we need not know exactly what "wire coral" and "cardinal fish" are in the example above, just their main features and their position in a type hierarchy of organisms. Second, the language for descriptive captions is often quite concrete, since it usually must describe real things and not abstrac-

tions, which means few verbs, and verbs are the hardest part of language understanding. Third, the forbidding-appearance specialized words in captions are generally nouns of grammatically simple subcategories (like the genus and species of organisms) that can rarely be confused with other English words. Fourth, software for interpreting restricted sublanguages has become better and more available recently.

Captions can do more than improve friendliness of a multimedia database system, however. They can actually speed access to multimedia data by providing additional, intelligent filtering of possible matches before retrieval. Thus caption-based access might well run faster than keyword-based, despite the greater overhead for query interpretation and more complex matching, because media data can often be large records retrieved from slow bulk storage. Furthermore, the user can interact with caption-based access to further improve it, by browsing through candidate captions and selecting good bets on a more informed basis than with keyword lists.

2. Previous work

Many researchers have worked on the problem of accessing multimedia data efficiently, although we know of no one who has tried to use captions in the central way that we do. Some research in information retrieval has investigated semantic representations of retrieval objects instead of keyword lists. The pioneering work of Kolodner (1983) embedded facts for retrieval in a complicated semantic network, and used a variety of special heuristics suggested by human reasoning to intelligently search that network. Cohen and Kjeldsen (1987) proposed spreading activation over a semantic network to find qualitatively good associative matches. Rau (1987) proposed a two-stage retrieval process from a semantic network, a spreading activation followed by graph matching; input questions (but not the data) were English, so much of the implementation was natural-language processing. Smith et al (1989) handled term-name differences between query and datum by using a hierarchy of concepts, where all levels could have pointers to retrieval objects. Sembok and van Rijsbergen (1990) translated natural-language texts into a predicate-calculus representation and then indexed terms for later retrieval.

Researchers in databases have been increasingly interested in multimedia databases. Some of this research concerns good ways of describing multimedia data for efficient retrieval, as the special summary data to describe pictures in Chang et al (1988) and the special parameters for describing video in Nagel (1988). Such descriptive information should be part of a good caption on the media datum. Other research concerns efficient administration of a database system containing multimedia objects, which can often be difficult because of its highly varied and highly storage-intensive formats. Bertino et al (1988), Roussopoulous et al (1988), Gibbs et al (1987), and Woelk et al (1986) exemplify this work, with an emphasis on conceptual modeling and query languages.

A longtime concern of artificial intelligence has been manipulating descriptions of the world, and many of its results apply to our problem. A variety of books address practical issues in knowledge representation, as Rowe (1988) and Davis (1990). Allen (1987) summarizes the state of the art in natural language processing. Grosz et al (1987) exemplifies the current state of natural-language processing tools, in presenting a powerful design tool for creating natural-language parsers and interpreters for a wide variety of domains. Katz (1988) has ideas about the special problem of using English for retrieval from databases.

An alternative to caption matching and indexing by keywords is content analysis of media data at query time, but this is usually too hard. There are some exceptions, such as scanning text to find a particular word. But such purely syntactic analysis is inflexible and of limited value for pictures, video, and audio for which inferencing is often needed. For instance, we could not match the fish picture to a query about life in coral forests, since coral is not visible in the picture. And additional information must always supplement content analysis, as for instance time of day or a picture's photographer.

3. Overview of our MARIE system

Fig. 1 shows a block diagram of the data structures in our MARIE system for efficient caption-based access to multimedia data, and Fig. 2 describes the blocks. MARIE is implemented in Quintus Prolog. At the top left in Fig. 1, human experts supply media data and their associated captions for storage in

DUDLEY KNOX LIBRARY
NAVAL POSTGRADUATE
MONTEREY CA 93943-510

the multimedia database, and at the top right, non-expert humans query the data. The media data (which comprise the *multimedia database*) are stored in a separate system on a separate processor, since they generally require much more space than the rest of the system. Pictures are the most common form of media data; each is at least the complexity of a television picture, so for a target of one million media data items, the multimedia database should be about 10^{11} bytes. This number and the generally read-only nature of the media data suggest optical storage. Our previous work of Meyer-Wegener et al (1989) and Holtkamp et al (1990) proposed details of management of the multimedia database, which we do not have space to discuss here.

The main innovation of our design is the access to media data through *meaning lists*, parsed and interpreted captions, instead of keywords. Meaning lists contain predicate-calculus expressions, and are equivalent to semantic networks; Fig. 3 gives an example. Meaning lists specify the meaning of each part of a natural-language utterance, then usually require that the conjunction of all meaning parts must hold. MARIE translates both English captions and English queries into meaning lists, the former in advance and the latter at query time.

Besides the captions themselves, MARIE requires auxiliary information from a lexicon, a concept hierarchy for the domain, and frame recognition rules. The *lexicon* (or dictionary) is necessary for parsing, and gives for each possible English word its part of speech, its grammatical forms, and the logical expression that represents it. The *concept hierarchy* is a type hierarchy on the possible concepts in meaning lists. It has both upward pointers (for semantic checking after parsing) and downward pointers (for finding captions with terms that are subtypes of those in the query); there can be more than one upward pointer from a concept. Lastly, the *frame-recognition rules* add inferences (usually generalizations) beyond what the natural language actually said.

The *coarse-grain search* does hash-table lookup of all occurrences of certain helpfully restrictive terms in the literals. This gives *caption pointers* to caption objects containing these terms, candidates for satisfying the query. Then the *fine-grain search* tries to match the full query meaning list against the candidate captions' meaning lists, binding variables as necessary.

A million media data items means a million captions. Judging from samples, the average caption will take 100 bytes: captions should summarize, not exhaustively catalog. So the caption database will be about 100 megabytes uncompressed, though compression techniques could reduce this. Note in Fig. 1 that some of the caption database is allocated to *supercaptions*. These are captions that describe a a set of media data, eliminating some redundancy; Fig. 3 shows some example supercaption information. Supercaptions are an important part of our design, and are a more user-friendly way of modeling hierarchical structure in data than an index on keywords.

After some preliminary experiments with a simple parser and a simple retrieval scheme for some pictures about World War II, we are now applying MARIE to photographs at the Naval Center. Eventually we intend to have 36,000 photographs and their captions online in an optical jukebox. Fig. 4 shows an example Sun-3 screen image from the current implementation. The query was "missile on an aircraft over a range", specified in the window at the lower right, and two small pictures were retrieved along with their registration information, shown in the lower left and lower middle of the screen; the upper right window shows parse-process information. (The pictures look better in color.)

4. Knowledge representation

With methodology and software developed in Rowe (1988), we put meaning lists in Prolog linked-list format, lists of literals expressing properties or binary relationships. To simplify matching, we limit predicates to a small set of primitive properties and relationships; for instance, we do not distinguish between "within", "inside", "part-of", "containing", and "comprising" relationships. However, we take care to represent the correct direction of relationships and to cover all words of the English input.

Conceptual generalization on the contents of meaning lists enables captions and queries to be considerably more informative. There are three kinds. First, a complete and thorough type hierarchy for the concepts (nouns and verbs) in the domain of discourse must be created. For instance for pictures of organisms, part is a species taxonomy, part is a taxonomy of observable characteristics of single organisms, part is a taxonomy of social characteristics, and part is a taxonomy of photographic terms. Type

information can be obtained from domain experts using techniques of knowledge acquisition for expert systems. Much of it can come from a natural-language dictionary, and it would be necessary anyway for finding subtypes of keywords, without which user-friendly access through keywords is impossible. It can be stored in the lexicon, since it helps determine the sense of verbs. Fig. 5 shows some lexicon entries from the 1951-word lexicon we used for the experiments reported in the last section of this paper; these are hashed and retrieved automatically by the Prolog interpreter.

A second kind of generalization information we use is the "frame" or "script" abstraction that frequently occurs in describing stereotypical human activities. "Coral", "fish", and "camera" in the cardinal-fish caption of section 1 suggest an observational underwater-photography activity using scuba gear; no single word indicates this, only the combination of clues. This is a "frame" or "script" problem and needs techniques like those in Schank and Abelson (1977). Such abstractions and their clues are usually highly topic-dependent, and must be obtained from an expert on the topic; they can be defined by rules that insert new terms into the lists, extra terms to exploit in matching. Our current implementation has some such rules in the final phase of meaning-list construction, and they are expressed as Prolog rules, but we could implement more.

A third kind of conceptual generalization is an idea previously not much explored: the *supercaption*, a caption that describes more than one media datum. For instance, the cardinal-fish caption could be a subcaption for the supercaption "Dive on 10/12/89 in Suruga Bay", which in turn could be a subcaption of the supercaption "1989 NGS/Tokyo Broadcasting System/Toba Aquarium project on Suruga Bay, Japan." A supercaption should be a full caption, not just a conceptual generalization like "dives". The Naval Air Warfare Center photographs have many supercaptions, often corresponding to tests conducted. Supercaptions can be obtained from a domain expert just like captions, and are most useful when they give information unobtainable from the concept hierarchy, like the dates, times, and places of a set of photos taken together. Supercaptions can create a hierarchy different from the type hierarchy; they can represent how an expert clusters media data using complex tradeoffs. "Registration" data, about how media objects were created, is often best expressed with supercaptions. For instance for a

photograph, this includes the photographer, the type of film, the exposure, the date and time the picture was taken, the place where the picture was taken, information that would require tedious labor to enter for every picture.

Our implementational approach to supercaptions is simple: we append all supercaptions (searching upward in the supercaption hierarchy) to the front of its subcaption to get the full subcaption for parsing, putting periods after the subcaption and supercaption if none were there before. That is, we assume additive semantics, and this works fine for nearly all supercaptions because our parser handles multi-sentence captions. This appending can be done when the database is entered, so its efficiency is not very important.

5. More about the natural-language understanding

We expect that most of the description of a media datum is best input in natural language. Other sources of descriptive information can supplement the natural language, like formatted registration data and any results of content analysis.

An illustration that the problem of understanding media-descriptive captions is considerably simpler than general natural-language understanding is provided by the statistics on the 31,000 distinct words from the 36,000 Naval Center picture captions (15,000 of which are codes and abbreviations), which we believe are typical of applications in which captions describe technical subjects and activities. Fig. 6 gives the frequencies of the 100 most common words among the 600,000 words of those captions. Most are nouns, and those that can be verbs can also be nouns (and do occur in the captions primarily as nouns). And the semantics of these words is relatively straightforward, except for the prepositions of which there are few in English. Thus a primary objective is a good type hierarchy for nouns.

Currently we are using the software DBG from Language Systems Inc. (Woodland Hills, California) for about half of our natural-language understanding component; we found its speed was reasonable on test sentences. We supply the lexicon, including the type information discussed in section 4, case infor-

mation, and morphology.

6. Query processing

We use a query-processing approach influenced by Rau's SCISOR (1987), with an emphasis on a variety of knowledge for different purposes; it used a two-phase search process.

6.1. Fine-grain search

We first find captions whose meaning lists match key terms of the query meaning list (*coarse-grain search*); then for each that matches the whole caption, we retrieve the corresponding media object (*fine-grain search*). Fine-grain search thus requires a subgraph-matching algorithm to match a caption to a query by binding variables and backtracking as necessary. Subgraph matching is much addressed in computer science, and there are algorithms for many special cases of it. In the worst case, the general subgraph-matching problem is exponential in complexity since the general algorithms are NP-hard. But the worst case will not likely to happen in real databases with real user queries, as it requires a single predicate name be used. We exploited the automatic backtracking features of the Prolog language in implementing the fine-grain matching.

6.2. Coarse-grain search

To handle our planned one million data items, we allocate $\log_2 10^6$=20 bits for each pointer. Judging from analysis of sample captions, there are about 20 indexable items per caption, 50 to be safe, so we need about 125 megabytes total for pointers from query terms to captions. This suggests the pointers be in secondary storage. Hashing to them is the simplest and fastest access method. So we identify key terms (which we define as nouns and verbs) in the meaning list translation of a user query, hash these to a secondary-storage table of caption pointers, intersect the pointer lists, and look up the corresponding captions. Partial matching can be permitted by a match threshold K, which is the number of lists intersected that must contain a pointer for the pointer to be considered acceptable.

Our hash table stores only exact matches. For instance, if a caption mentions cardinal fish, then only the hash table entry for "cardinal fish" points to it, not the entry for "fish". So a query that mentions just "fish" must use the concept hierarchy to reach other hash-table entries to find the cardinal-fish caption. This saves much space at the expense of (main-memory) time to follow the downward pointers. We also save space by using supercaption pointers in the hash table.

Disjunctions are treated just like the subtypes and subcaptions, which are implicit disjunctions. (Disjunctions in captions should be usually rejected as too vague to be a good description.) Also, other kinds of inheritance besides the type inheritance of section 4 can be exploited (Rowe (1988), Rowe (1991)). For instance, a query asking for pictures of planes with ceramic-composite wings should match a ceramic-composite plane, since a wing is part of a plane. This kind of inference won't work at all for certain properties (like cost) and works in the opposite direction for other properties (like defectiveness of a part, which inherits upwards to give defectiveness of a plane containing the part). A rule-based inference system covers the cases; the last entry in Fig. 5 illustrates the word-specific information necessary for such rules.

Once pointers to media data have been found, it is often cost-effective to retrieve only the captions first. Then users may be able to rule out some of them without an expensive media datum fetch, and such selections also provide relevance feedback for future partial matches.

7. Experimental results

To test our implementation, we randomly selected 217 images and associated captions from the Photo Lab (the photographic archive) of the Naval Air Warfare Center. The captions totalled 4488 words, from which we built a 1951-word lexicon (including some words from an earlier application) and a 830-word type hierarchy on nouns and verbs. Then we asked Photo Lab personnel to provide us with typical queries asked them; they supplied us with 46, 2 of which involved concepts not in captions. We ran MARIE on the 44 remaining queries, averaging 4.9 words in length; mean processing time was 14.1 seconds of CPU time and the median was 4.2 seconds, with 2 queries needing to be rephrased because

of parse failure. No concurrent processing was used. We then had Photo Lab personnel judge the acceptability (yes/no) of the computer-selected photographs. From these tests, without changing the natural-language processor, we had a recall of 93.6% and a precision of 94.7%, which suggest soundness of the implementation. Photo Lab personnel also agreed our system was very easy to use. More details are in Guglielmo (1992).

8. References

Allen, J. (1987). *Natural language understanding*. Menlo Park, CA: Benjamin Cummings.

Bertino, E., Rabitti, F., and Gibbs, S. (1988, January). Query Processing in a Multimedia Document System. *ACM Transactions on Office Information Systems, 6*, 1, 1-41.

Chang, S. K., Yan, C. W., Dimitroff, D. C., Arndt, T. (1988, May). An Intelligent Image Database System. *IEEE Transactions on Software Engineering, 14*, 5, 681-688.

Cohen, P. R. and Kjeldsen, R. (1987). Information retrieval by constrained spreading activation in semantic networks. *Information Processing and Management, 23*, 4, 255-268.

Davis, E. (1990). *Representations of commonsense knowledge*. Palo Alto, CA: Morgan Kaufmann.

Gibbs, S., Tsichritzis, D., Fitas, A., Konstantas, D., and Yeorgoroudakis, Y. (1987). Muse: A multimedia filing system. *IEEE Software, 4*, (2), 4-15.

Grosz, B., Appelt, D., Martin, P. and Pereira, F. (1987). TEAM: An experiment in the design of transportable natural language interfaces. *Artificial Intelligence, 32*, 173-243.

Guglielmo, E. (1992, August). Intelligent Information Retrieval for Multimedia Databases Using Captions. Ph.D. thesis, Department of Computer Science, U.S. Naval Postgraduate School, Monterey, California USA.

Holtkamp, B., Lum, V., and Rowe, N. (1990, October). DEMOM--A description-based media object

data model. Proceedings of the IEEE Computer Software and Applications Conference (COMPSAC), Chicago IL.

Katz, B. (1988, March). Using English for indexing and retrieval. Proceedings of RIAO-88, Cambridge, MA, 314-322.

Kolodner, J. (1983, September). Indexing and retrieval strategies for natural language fact retrieval. *ACM Transactions on Database Systems, 8*, 3, 434-464.

Meyer-Wegener, K., Lum, V., and Wu, C. (1989). Image database management in a multimedia system. In *Visual database systems*,IFIP TC 2/WG 2.6 Working Conference, Tokyo, Japan, ed. T. Kunii, North Holland, Amsterdam, 497-523.

Nagel, H. (1988, May). From image sequences towards conceptual descriptions. *Image and Vision Computing, 6*, 2, 59-74.

Rau, L. (1987). Knowledge organization and access in a conceptual information system. *Information Processing and Management, 23*, 4, 269-284.

Roussopoulous, N., Faloutsos, C., and Sellis, T. (1988, May). An efficient pictorial database system for PSQL. *IEEE Transactions on Software Engineering, 14*, 5 639-650.

Rowe, N. (1988). *Artificial Intelligence through Prolog*. Englewood Cliffs, N.J.: Prentice-Hall.

Rowe, N. (1991). Management of regression-model data. *Data and Knowledge Engineering, 6*,349-363.

Schank, R. and Abelson, R. (1977). *Scripts, plans, goals, and understanding*. Hillsdale, N.J.: Lawrence Erlbaum.

Sembok, T. and van Rijsbergen, C. (1990). SILOL: A simple logical-linguistic document retrieval system. *Information Processing and Management, 26* (1), 111-134.

Smith, P., Shute, S., Galdes, D., and Chignell, M. (1989, July). Knowledge-based search tactics for an intelligent intermediary system. *ACM Transactions on Information Systems, 7*, 3. 246-270.

Woelk, D., Kim, W., and Luther, W. (1986, May). An object-oriented approach to multimedia databases. Proceedings of ACM SIGMOD 86 International Conference on Management of Data, Washington, DC, 311-325.

REPRODUCED AT GOVERNMENT EXPENSE

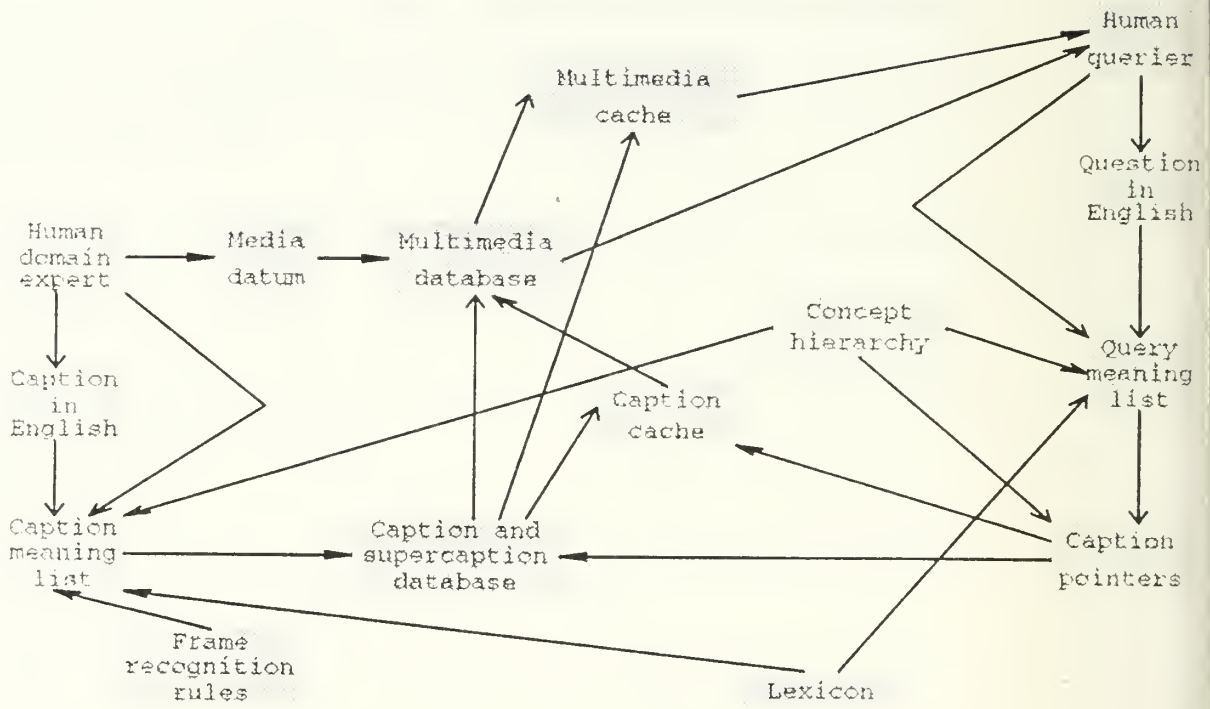

Figure 1: Block diagram of our MARIE system.

REPRODUCED AT GOVERNMENT EXPENSE

Structure	*Description*	*Megabytes*	*Storage type*
lexicon	language dictionary	1	main memory
concept hierarchy	complete type info	0.1	main memory
frame-recognition rules	recognizes plans in captions	0.1	main memory
hash table to caps., supercaps.	maps from query terms	100	magnetic disk
caption database	meaning lists	100	secondary storage
caption cache	most recent ones	1	magnetic disk
multimedia database	the actual data	>100,000	optical jukebox
multimedia cache	most recent ones	100	magnetic disk

Figure 2: Data structures, with approximate sizes, for a million-object multimedia database with media datum items at least 10000K bytes each.

REPRODUCED AT GOVERNMENT EXPENSE

Caption: "Sidewinder AIM 9R missile mounted on F/A-18C BU# 163284 aircraft, nose 110. Closeup view of front of missile and launcher."

Frame inferred: equipment-description

Example meaning terms inheritable from supercaptions: [photograph(color), focus(medium-range)]

Meaning list (actual parser output):
theme('pastpart(262870-1-1)',obj('noun(262870-1-3)')).
event('pastpart(262870-1-1)',rise).
ref_pt('noun(262870-1-3)',front).
loc('noun(262870-1-3)',on('noun(262870-1-6)')).
inst('noun(262870-1-3)','AIM 9R').
inst('noun(262870-1-6)','F/A-18C').
ref_pt('noun(262870-2-3)',front).
inst('noun(262870-2-3)',launcher).
tag('noun(262870-1-7)',id_of('noun(262870-1-6)')).
mods('noun(262870-1-7)',designator('110')).
inst('noun(262870-1-7)',nose).
theme('noun(262870-2-1)',of('noun(262870-1-3)')).
theme('noun(262870-2-1)',of('noun(262870-2-3)')).
mods('noun(262870-2-1)',quant(closeup)).
inst('noun(262870-2-1)',view).
tag('noun(262870-1-5)',id_of('noun(262870-1-6)')).
mods('noun(262870-1-5)',designator('163284')).
inst('noun(262870-1-5)',bureau_no).

Figure 3: An example caption and corresponding meaning list output from the current MARIE system, plus examples of additional information inferrable or inheritable. Note: hyphenated terms refer to caption words; e.g., "noun(262870-1-5)" means the fifth word of the the first sentence for photo 262870.

REPRODUCED AT GOVERNMENT EXPENSE

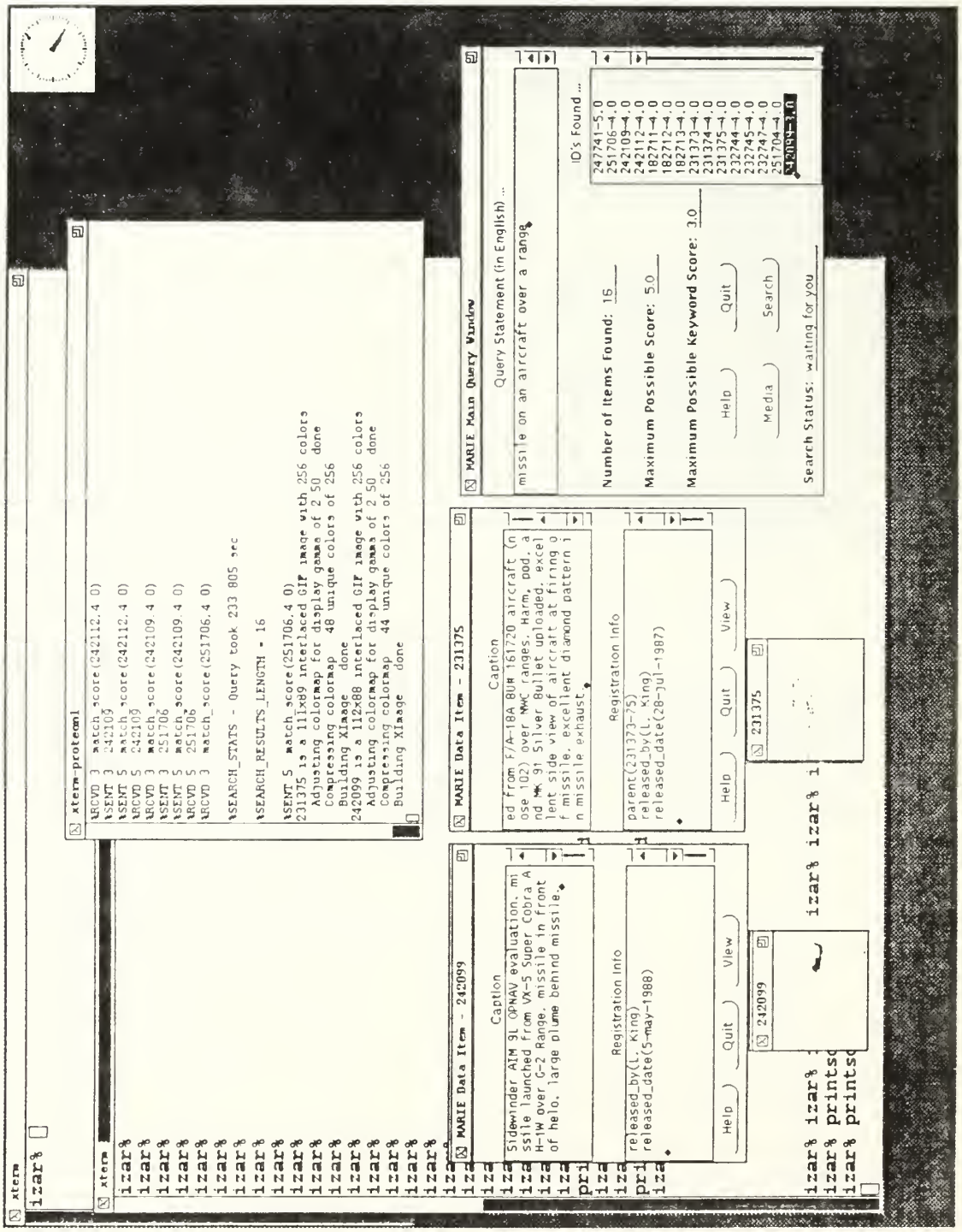

Figure 4: An example picture of the workstation screen while running MARIE.

REPRODUCED AT GOVERNMENT EXPENSE

"Sidewinder" is a noun of syntactic type 9 (a proper noun that can have articles in front of it), must be capitalized, and is a kind of missile:
noun(`Sidewinder`,morph(9),fp(missile)).

"Missile" is a noun of syntactic type 1 (a common noun whose plurals are formed by adding "s"), and is a kind of physical object:
noun(missile,morph(1),fp(phys_obj)).

"Impact" is a verb of syntactic type 1-a (a verb whose third person singular ends in "s", whose past participle ends in "ed", and whose present participle ends in "ing"), its synonym is "hit", and its direct object must be a physical object:
verb(impact,morph(1-a),fpcat(hit),case([[dobj(phys_obj)]]).

Any missile, when the word is used in the most common sense of the term, has a bulkhead, Dev-Assist, dome, engine, homing device, tail fin, warhead, and TDD; and a missile is always part of an attack aircraft:
slot(missile,noun-1,correlations,
[c(has_part,bulkhead), c(has_part,`Dev-Assist`), c(has_part,dome),
c(has_part,engine), c(has_part,`homing device`), c(has_part, `tail fin`),
c(has_part,warhead), c(has_part,`TDD`), c(part_of,`attack aircraft`)]).

Figure 5: Example entries in the current lexicon of MARIE, preceded by their interpretations. Note the first three include type hierarchy information. The fourth includes part-whole relationships necessary for inferences.

REPRODUCED AT GOVERNMENT EXPENSE

and (17790)	test (14160)	of (14012)	view (11043)	on (9821)
in (8172)	with (7964)	at (6149)	aircraft (6059)	to (5437)
views (4701)	from (3601)	missile (3472)	post (3384)	bldg (3301)
sled (3277)	firing (3207)	air (3136)	aerial (3040)	pre (2865)
front (2807)	side (2672)	oblique (2648)	1 (2627)	released (2548)
for (2521)	looking (2499)	the (2474)	mk (2473)	excellent (2326)
lab (2291)	target (2283)	range (2240)	run (2191)	warhead (1988)
showing (1981)	motor (1953)	cookoff (1935)	facility (1851)	launcher (1814)
2 (1781)	personnel (1775)	area (1751)	sidewinder (1718)	lake (1692)
bomb (1671)	center (1565)	tail (1564)	rocket (1549)	track (1523)
closeup (1515)	3 (1512)	copy (1481)	overall (1470)	studio (1367)
right (1358)	program (1329)	3/4 (1320)	by (1306)	inch (1289)
fuze (1286)	left (1281)	various (1275)	graphics (1270)	0 (1268)
before (1266)	s (1264)	sn (1261)	a (1259)	control (1234)
china (1206)	5 (1179)	nwc (1162)	system (1120)	l. king (1111)
rear (1102)	fast (1100)	after (1098)	mod (1095)	background (1091)
michelson (1064)	vertical (1053)	seat (1046)	full (1043)	launch (1001)
ejection (970)	flight (964)	facilities (935)	ii (931)	over (921)
site (920)	n (913)	asroc (911)	x (906)	npc (899)
aim (891)	portrait (865)	north (850)	construction (831)	dummy (827)

Figure 6: The 100 most frequent words in 36,000 captions (600,000 words) for the Naval Weapons Center photographic database, with their frequencies.

Distribution List

SPAWAR-3242
Attn: Phil Andrews
Washington, DC 20363-5100 1

Defense Technical Information Center,
Cameron Station,
Alexandria, VA 22314 2

Dudley Knox Library, Code 0142,
Naval Postgraduate School,
Monterey, CA 93943 2

Center for Naval Analyses
4401 Ford Ave.
Alexandria, VA 22303-0268 1

Research Office
Code 08
Naval Postgraduate School,
Monterey, CA 93943 1

John Maynard
Code 402
Command and Control Departments
Naval Ocean Systems Center
San Diego, CA 92152 1

Dr. Sherman Gee
ONT-221
Chief of Naval Research
800 N. Quincy Street
Arlington, VA 2217-5000 1

Leah Wong
Code 443
Command and Control Departments
Naval Ocean Systems Center
San Diego, CA 92152 1

Bernhard Holtkamp
University of Dortmund
Dept. of Computer Science
Software-Technology
P.O. Box 500 500
D-4600 Dortmund 50
West Germany 5

Vincent Y. Lum
Code CSLu
Naval Postgraduate School
Monterey, CA 93943 5

Dr. Neil C. Rowe, Code CSRp
Computer Science Department
Monterey, CA 93943 20

Klaus Meyer-Wegener
University of Kaiserslautern
Computer Science Department
P.O. Box 30 49
D-6750 Kaiserslautern
West Germany 1

Professor Robert B. McGhee, Code CSMz
Department of Computer Science
Naval Postgraduate School
Monterey, CA 93943 1

DUDLEY KNOX LIBRARY
3 2768 00347443 8